

A Stigmergy Based Approach to Data Mining

Manu De Backer¹, Raf Haesen¹, David Martens¹, and Bart Baesens^{1,2}

¹ Department of Applied Economic Sciences, K.U.Leuven, Belgium,
Naamsestraat 69, B-3000 Leuven, Belgium

{Manu.Debacker, Raf.Haesen, David.Martens, Bart.Baesens}@econ.kuleuven.be

² University of Southampton, School of Management, United Kingdom,
Highfield Southampton, SO17 1BJ, United Kingdom

Bart@soton.ac.uk

Abstract. In this paper, we report on the use of ant systems in the data mining field capable of extracting comprehensible classifiers from data. The ant system used is a *MAX-MIN* ant system which differs from the originally proposed ant systems in its ability to explore bigger parts of the solution space, yielding better performing rules. Furthermore, we are able to include intervals in the rules resulting in less and shorter rules. Our experiments show a significant improvement of the performance both in accuracy and comprehensibility, compared to previous data mining techniques based on ant systems and other state-of-the-art classification techniques.

1 Introduction

In this paper we focus our attention on the use of an ant system capable of extracting comprehensible, accurate rules from categorical data that is competitive with state-of-the art classification techniques. Artificial ant systems are inspired on the behavior of real ant systems. The general idea of the system is the following [1]: a number of computational concurrent and asynchronous agents (ants) move through their environment and by doing so incrementally construct a solution for the problem at hand. Ants move by applying a stochastic local decision policy based on two parameters, called the trail value (pheromone value) and a problem dependent value (heuristic value). Ant systems have shown to be a viable method for attacking hard combinatorial optimization problems. Stützle et al. [6] advocate that improved performance can be obtained by a stronger exploitation of the best solutions. According to them, the key to achieve best performance is to combine an improved exploitation of the best solutions found with an effective mechanism for avoiding early search stagnation. A *MAX-MIN* ant system (*MMAS*) differs from a normal ant system in three aspects. First of all, after each iteration only the best ant adds pheromone to its trail. Secondly, the range of possible pheromone trails on each solution component is limited to an interval $[\tau_{min}, \tau_{max}]$. Finally, each trail is initialized with a pheromone value of τ_{max} .

2 AntMiner+

The first application of ant systems for data mining was reported in [4], where the authors described the AntMiner algorithm for the discovery of classification rules. Extensions and optimizations of the AntMiner are described in AntMiner2 and AntMiner3 [3]. The aim is to extract simple **if-then-else** rules from data. We build further on the work introduced before and try to resolve some issues. First of all, we define the environment as a directed, acyclic construction graph which allows a clear representation of the problem domain and considerably improves the performance of the ant system. Furthermore, we introduce the better performing *MAX-MIN* ant system for mining rules.

The construction graph is defined as follows: each 'column' or node group corresponds to a variable and every 'row' corresponds to a value. Each ant going from node $n_{i,j}$ to node $n_{i+1,k}$ adds the term $\langle V_{i+1} = Value_k \rangle$ to its rule. To allow for rules where not all the variables are involved, an extra *dummy node* is added whose value is undetermined, meaning it can take any of the values available. Although only categorical variables can be used in our implementation, we make a distinction between nominal and ordinal variables. Each nominal variable has one node group, but for the ordinal however, we build two node groups to allow for intervals to be chosen by the ants. The first node group corresponds to the lower bound of the interval and should thus be interpreted as $\langle V_{i+1} \geq Value_k \rangle$, the second node group determines the upper bound, giving $\langle V_{i+2} \leq Value_k \rangle$. This allows to have less, shorter and actually better rules. Figure 1 gives a general view of the construction graph with the first variable being nominal and the second one ordinal, hence having two node groups.

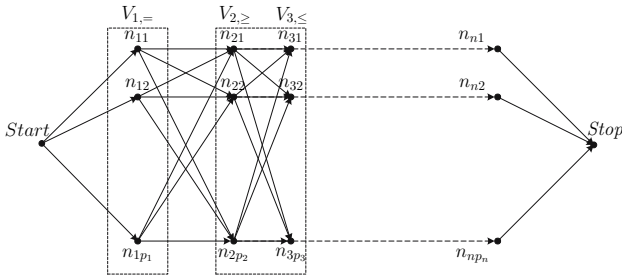


Fig. 1. AntMiner+ construction graph

The edge probability $P_{(n_{i,j}, n_{i+1,k})}$ is the probability that an ant which is in node $n_{i,j}$ (node for which variable V_i is equal to its j^{th} value) will go to node $n_{i+1,k}$. This probability is defined as follows:

$$P_{(n_{i,j}, n_{i+1,k})} = \frac{[\tau_{(n_{i,j}, n_{i+1,k})}]^\alpha \cdot [\eta_{n_{i+1,k}}]^\beta}{|V_{i+1}| \sum_{l=1} [\tau_{(n_{i,j}, n_{i+1,l})}]^\alpha \cdot [\eta_{n_{i+1,l}}]^\beta} \quad (1)$$

Notice that this probability is dependent on two values: the heuristic value and the pheromone value. The importance of these values is determined by the α and β parameters. The heuristic value gives for each node a notion of its importance in the problem domain. For data mining, the importance is measured by the number of data points that are covered (described) by a value. Since we extract rules for the $class = 1$ case, we define the heuristic value for node $n_{i,k}$ as follows:

$$\eta_{n_{i,k}} = \frac{|V_i = Value_k \ \& \ class = 1|}{|V_i = Value_k|} \quad (2)$$

Generally, updating the pheromone trail of an ant system is accomplished in two phases, viz evaporation and reinforcement. Applying the ideas of $\mathcal{MAX-MIN}$ ant systems has direct consequences for the pheromone updating rule. First of all, evaporation in an ant system is accomplished by diminishing the pheromone level of each trail according to the following rule, with ρ the evaporation factor:

$$\tau_{(n_{i,j}, n_{i+1,k})}(t+1) = \rho \cdot \tau_{(n_{i,j}, n_{i+1,k})}(t) \quad (3)$$

Secondly, in a $\mathcal{MAX-MIN}$ ant system, reinforcement of the pheromone trail is only applied to the best ant's path. Taking into account the evaporation factor as well, the update rule for the best ant's path can be described as:

$$\tau_{(n_{i,j}, n_{i+1,k})}(t+1) = \rho \cdot \tau_{(n_{i,j}, n_{i+1,k})}(t) + \Delta^{best} \quad (4)$$

Clearly, the reinforcement of the best ant's path should be proportional to the quality of the path. For data mining we define the quality of a rule by the sum of its *confidence* and its *coverage*, as defined by equation 5. Confidence is an indication of the number of correctly classified data points by a rule compared to the total number of data points covered by that rule. The coverage gives an indication of the overall importance of the specific rule by looking at the number of correctly classified data points over the total number of data points.

$$\Delta^{best} = \underbrace{\frac{|rule_{best-ant} \ \& \ Class = 1|}{|rule_{best-ant}|}}_{\text{confidence}} + \underbrace{\frac{|rule_{best-ant} \ \& \ Class = 1|}{|Cov = 0|}}_{\text{coverage}} \quad (5)$$

An additional restriction imposed by the $\mathcal{MAX-MIN}$ ant systems is that the pheromone level of the edges is restricted by an upper-bound (τ_{max}) and a lower-bound (τ_{min}).

3 Experiments and Results

We applied AntMiner+ to several publicly available datasets. We conducted 10 runs for each dataset where the data is randomized and the first $\frac{2}{3}$ of the data is taken as training data and the remaining $\frac{1}{3}$ as test data. Experiments are conducted on the tic-tac-toe and Breast Cancer Wisconsin datasets [2], and Ripley's dataset [5]. The results are shown in Table 1, also included are the

Table 1. Average out-of-sample performance

	tic-tac-toe		BCW		Ripley	
	inst	attr	inst	attr	inst	attr
	958	9	699	9	1250	2
Technique	Acc	#R	Acc	#R	Acc	#R
AntMiner+	99.75	8	95.58	1	89.41	3.9
AntMiner1	70.99	16.5	92.63	10.1		
AntMiner3	76.58	18.6	94.32	13.2		
C4.5	76.96	23	94.38	6	89.04	6

results for C4.5 and the results published for the tic-tac-toe and Breast Cancer Wisconsin datasets from AntMiner1 and AntMiner3 [3]. For each dataset, the number of instances (inst) and attributes (attr) as well as the accuracy and number of generated rules are displayed. The best performance measure for each dataset is shown in boldface.

The better results of AntMiner+ can be attributed to our $\mathcal{MAX}\text{-}\mathcal{MIN}$ approach, our simple construction graph with the inclusion of dummy nodes, as well as our ability to include intervals in our rules.

4 Conclusion

Ant systems are a nature inspired technique where ants communicate through the principle of stigmergy. Although ants have a limited memory and perform actions based on local information only, the ants are able to come to complex behavior due to self-organization and indirect communication. We defined our environment as a simple, though complete construction graph so that the ants can construct rules while going from source to sink. The solution provided by AntMiner+ provides comprehensible, accurate classifiers and performs better than or competitive with state-of-the-art classification techniques.

References

1. M. Dorigo, V. Maniezzo, and A. Colorni. Positive feedback as a search strategy. Technical Report 91016, Dipartimento di Elettronica e Informatica, Politecnico di Milano, IT, 1991.
2. S. Hettich and S. D. Bay. The uci kdd archive [<http://kdd.ics.uci.edu>], 1996.
3. B. Liu, H. A. Abbass, and B. McKay. Classification rule discovery with ant colony optimization. *IEEE Computational Intelligence Bulletin*, 3(1):31–35, 2004.
4. R. S. Parpinelli, H. S. Lopes, and A. A. Freitas. An ant colony based system for data mining: Applications to medical data. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 791–797, San Francisco, California, USA, 7-11 2001. Morgan Kaufmann.
5. B. D. Ripley. Neural networks and related methods for classification. *J. Roy. Statist. Soc. B*, 56:409–456, 1994.
6. T. Stützle and H. Hoos. $\mathcal{MAX}\text{-}\mathcal{MIN}$ ant system. *Future Generation Computer Systems*, 16(8):889–914, 2000.